Archives of Computational Methods in Engineering (2024) 31:1701–1719
https://doi.org/10.1007/s11831-023-10028-9

REVIEW ARTICLE

Advances in Deep Learning Models for Resolving Medical Image Segmentation Data Scarcity Problem: A Topical Review

Ashwini Kumar Upadhyay[1,2] · Ashish Kumar Bhandari[1]

Received: 5 March 2023 / Accepted: 26 October 2023 / Published online: 9 December 2023

Abstract
Deep learning (DL) methods have recently become state-of-the-art in most automated medical image segmentation tasks. Some of the biggest challenges in this field are related to datasets. This paper aims to review the recent developments in deep learning architectures and approaches that aim to resolve dataset-related challenges faced in DL-based medical image segmentation. We have studied architectural developments in deep learning models and their recent applications in medical image segmentation tasks. Popular U-Net-based models are tested for segmentation performance comparison on a Coronavirus disease 2019 (Covid-19) lung infection Computed Tomography segmentation dataset. The comparison results prove the effectiveness of the original U-Net architecture, even in present-day medical image segmentation tasks. To overcome major dataset-related challenges such as labeled data scarcity, high annotation time and cost, distribution shifts, low-quality of images, and generalizability issues; we have studied recent developments in deep learning approaches like active learning, data augmentation, domain adaptation, and self- and semi-supervised learning, that aim to provide innovative solutions for those challenges. With rapid developments in the field, approaches like data augmentation, domain adaptation, and semi-supervised learning have become some of the hot areas of research, aiming for more efficient use of datasets, better segmentation prediction, and model generalizability.

1 Introduction

Medical image diagnosis through automatic image analysis has been revolutionized by the recent advances in computer vision (CV) with deep learning models. Image segmentation is a major process step in object detection [1] and in the segmentation process, an image is partitioned into different regions based on certain features in that image [2]. The objective of segmentation is to predict the class label of each pixel, grounded on the region of interest (ROI) in an image, within which that pixel is situated [3]. Medical image segmentation (MedSeg) is an important step in medical image analysis in which it aims to identify pixels associated with specific organs or lesions in medical images such as the chest, brain, abdomen, eye, heart, etc. [4]. MedSeg processes are of 3 types: (1) Manual segmentation by the radiologist (used to produce labeled image data for deep learning models). Manual segmentation is time-taking, costly, and expert-dependent. (2) Fully automated segmentation methods, which uses computational models to produce results without human intervention. One of the biggest challenges for learning-based automated MedSeg methods are the scarcity of labeled data, images with high variabilities, distribution shifts, noises, and low contrast [5]. (3) Semi-automatic segmentation is the combination of the above two methods [6].

Earlier, MedSeg approaches used template matching, active contours, statistical and edge detection models along with machine learning (ML) techniques [7–12]. Before the popularity of deep learning (DL) structures, academics used several mathematical models [13] and conventional ML techniques [14] for segmenting medical images. But with the rapid progress in the DL approaches, artificial neural network (ANN) based structures, especially models based on convolutional neural networks (CNN) have become one of the hottest areas of research for CV tasks such as image segmentation. CNN-based structures give some of the

✉ Ashish Kumar Bhandari
bhandari.iiitj@gmail.com

Ashwini Kumar Upadhyay
ashwiniu.ph21.ec@nitp.ac.in

1 Department of Electronics and Communication Engineering, National Institute of Technology Patna, Patna 800005, India

2 Department of Electronics Engineering, Rajkiya Engineering College Kannauj, Tirwaganj 209732, India

best MedSeg outcomes as they generally are not affected by image quality-related factors such as moderate levels of noise and contrast issues [15]. DL-based techniques have consistently been evolving and outperforming all the other techniques for a majority of the automated MedSeg tasks [16–22]. Depending on the type of datasets used for training, DL methods can be broadly divided into 3 types: (1) Supervised learning based models such as earlier CNN-based models which used only labeled data for model training, (2) Unsupervised learning based models such as generative adversarial networks (GAN) which can train on unlabeled datasets, and (3) Semi-supervised learning (SSL) based models which, essentially are a combination of supervised and unsupervised learning techniques [23]. With the advancement of technology, a huge amount of labeled image data is now available for supervised learning techniques but in the medical imaging domain, finding a large amount of labeled data is still a challenge due to data privacy issues, higher cost, and time required in the annotation process [5]. A purely supervised learning technique requires enormous amount of labeled data, whereas an SSL based technique can use large amount of available unannotated data with the help of little amount of available annotated data. Purely unsupervised learning techniques can detect abnormal areas, but they struggle to associate these regions with a target disease [24–26]. Data augmentation techniques aspire to enhance the size and quality of limited available labeled datasets which makes them potentially very useful in MedSeg [27]. It is for these reasons, most of the recent works in MedSeg have focused on using combinations of the above-mentioned techniques, rather than focusing on just one for solving dataset-related problems.

Previous works reviewing DL-based methods for MedSeg have mostly focused on either deep network structural developments [4, 28, 29] or methods based on a specific network architecture such as U-Net [30–32] or GAN [33–35]. For example, Taghanaki et al. [36] categorized papers based on works on architectural developments as well as works based on training-related hyperparameter selection such as loss function, data synthesis, and supervision techniques. Some studies such as the ones by Cheplygina et al. [37] and Tajbakhsh et al. [5], focused on methods that solve a specific problem such as the shortage of high-quality annotated medical image data for training. There have been studies that took a particular DL-based approach for review. For example, Bohlender et al. [38] reviewed DL-based methods that use some kind of shape-constraint mechanism for improving the segmentation performance of the model. Xie et al. [39] surveyed DL models that improved their segmentation performance by utilizing domain knowledge of a medical image modality provided by medical experts and Zhou et al. [40] studied methods that used multi-modality fusion techniques. Some other studies focused on models performing segmentation on particular organ(s) or for the detection of a particular disease; for example, Devunooru et al. [41] and Wadhwa et al. [42] reviewed works on brain tumor segmentation, Fu et al. [43] studied models focusing on multi-organ segmentation, and Shi et al. [44] included works on the task of segmentation of lung infection for Covid-19 detection.

For initial literature collection, we searched with the keywords "deep learning", "CNN", "U-Net", "segmentation" and "medical image" on PubMed and Google Scholar. Subsequently, we manually compiled the papers we found and carefully examined their abstracts to identify suitable methods that address challenges associated with datasets in the context of deep learning-based medical image segmentation. Lastly, we thoroughly reviewed the references and citations of the chosen papers to identify additional pertinent works. The majority of the papers we collected were published in 2018 or later, enabling us to focus on recent advancements in deep learning architectures and approaches, including active learning, data augmentation, domain adaptation, and self- and semi-supervised learning, all of which are relevant to medical image segmentation.

Unlike the previous reviews, we have not only focused on the architectural developments in deep learning, but also on the developments in different non-architectural approaches used for solving major dataset-related challenges plaguing DL-based models in MedSeg tasks. By incorporating DL based studies on diverse set of modalities, organs, and challenges in this work, we can recognize broad directions of research in the larger domain of DL based MedSeg. Additionally, some popular U-Net-based models are tested on a 'Covid-19 lung infection segmentation CT dataset' [45] to prove the effectiveness of original U-Net architecture even in present-day MedSeg tasks.

This paper has been divided into the following three sections:

(i) Deep learning architectures for MedSeg: in this section we will discuss the development of convolutional network architectures from CNN to fully convolutional networks (FCN), U-Net, and GAN along with their use in MedSeg, broadly aiming for efficient and accurate segmentation with medical dataset constraints. We have also implemented some of the popular U-Net-based models on a chest CT dataset to compare their segmentation performances.
(ii) Approaches for DL based MedSeg: Here, works related to the approaches developed for solving major dataset-related issues concerning DL-based MedSeg will be discussed. These approaches mostly focus on the non-architectural aspects of DL and try to resolve various problems plaguing DL-based MedSeg. For example, active learning techniques for improving annotation process efficiency, domain adaptation for

using cross-domain datasets, semi-supervised learning for utilizing unlabeled datasets, and data augmentation techniques for enlarging the limited training dataset, etc.

(iii) Challenges and opportunities: In this part, we will discuss the knowledge gained from this survey to identify the major challenges and their possible solutions. We will also discuss the hot areas of research in the field of DL-based MedSeg to guide future researchers.

2 Deep Learning Architectures for MedSeg

DL-based networks are a family of artificial neural networks (ANN) that mimic the way a human brain works, and fall under the larger family of machine learning (ML) approaches [46]. Unlike the conventional methods, DL-based networks can work directly on the raw images, without the need for any significant image pre-processing. DL models are deeper versions of conventional ANN as they contain a large number of ANN layers and network parameters [20]. Each layer in a DL model takes input from the previous layer and hierarchically represents its features. Generally, successive layers in the model learn about the edges, their position, and the combination of these edges in the form of patterns enabling the final layers to perform object detection [47]. Hence with these properties, DL-based models are increasingly becoming state-of-the-art in most of the computer vision applications recently [48]. Convolutional networks such as CNN are DL models, exclusively developed for CV tasks, and GANs are used for synthesizing images in different domains. In this section, we will be discussing the architectural evolution of convolutional networks and GANs for a variety of MedSeg applications.

2.1 Convolutional Neural Networks (CNN)

CNNs are very popularly used for CV tasks such as image classification and segmentation [49]. A typical CNN structure has multiple layers of convolution followed by activation function and pooling operators, with fully connected dense layers at the end. In the convolution operation, a filter (kernel) is convolved with the input image to produce feature representations of the image. Elements of the filter are learned at the time of training. The pooling layer reduces the input dimensionality by representing the image in a smaller size. It helps to avoid overfitting and get better generalizations. Activation function, such as rectified linear unit (ReLU) [50], defines the output of a neuron, given a set of inputs. The fully connected dense part of the network is used for final object classification. The parameters (weights and biases) and filter (kernel) values are learned and updated using learning (optimization) algorithms such as adaptive moment estimation (Adam) [51] with backpropagation during the network training to minimize the loss [52]. Given the capabilities of CNN in classification and object detection tasks, researchers have explored different types of CNN architectures for MedSeg tasks.

In a two-dimensional (2D) CNN structure, 2D filters are applied to 2D images for segmentation, and multiple 2D images in different channels (like RGB) can be given to a CNN input for an improved segmentation result [53]. A 2.5D CNN approach utilizes greater pixels neighborhood information where 2D filters are used on 2D image patches represented in the 3 planes of cartesian coordinates [54]. This approach has been used for the segmentation of multiple organs, segmenting cartilage of knee joints, etc. [55, 56]. 2D annotated image data is more easily accessible than three-dimensional (3D) images. Moreover, 3D (volumetric) images can be decomposed into 2D images, that can be used with 2D kernels in a 2.5D CNN structure [57]. Since 2.5D CNN structures use 2D kernels, the representations across the three axes (X, Y, & Z) are of lower quality, which can be improved by using a 3D CNN structure. Structures of a 3D CNN and 2D CNN are similar with the difference that 3D CNN has 3D kernels (convolution filter layers) and 3D pooling (down-sampling) layers [58]. Advancement of hardware and software technologies and the availability of 3D medical images like CT scans, Magnetic Resonance Imaging (MRI), etc. have encouraged researchers to utilize 3D spatial information in 3D images for better segmentation. 3D CNN structures have been used for brain boundary and tumor detection [59, 60]. The deeper a deep neural network is, the higher will be its capability to learn features. But deeper networks are very hard to train and suffer from problems like vanishing gradients and degradation. A residual learning framework where a residual feature map is fed to every few layers (working as skip connections), can ease the process of training deeper neural networks [61]. Convolutional residual network (CRN) developed by He et al. [61] was used for natural image segmentation and achieved accuracies on par with deeper designs. VoxResNet, a type of 3D CRN developed by Chen et al. [62] was used to segment 3D brain MR images. The performance of VoxResNet in segmenting tissues from 3D brain MR images proves the efficacy of deeper models and also gives solution to the problem of degradation.

Although conventional CNN based models perform well in segmenting medical images if trained and tested on similar datasets, they are unable to produce accurate results in real-world clinical practices. This is because model generalizability and image-specific adaptations are some of the biggest challenges CNNs face in addition to the inherent problems with medical images such as poor image quality, variations in imaging, segmentation processes, and patients

[63]. There have been studies [64–66] in which CNNs have been used with user interactions where an expert user can intervene in the training process for improvements, which often lead to more robust segmentations [63]. For MedSeg, Wang et al. [67] proposed a CNN-based model with scribble and bounding box-based segmentation framework. In this model, image-specific fine-tuning was used that improve the model's generalizability. The framework was used for the 2D segmentation of organs from fetal magnetic resonance (MR) slices and 3D segmentation of brain tumors from different MR sequences. The framework with supervised (user interaction based) and unsupervised refinement methods produced good results on previously unseen objects.

2.2 Fully Convolutional Network (FCN)

The conventional CNN structure has some inherent problems like loss of spatial information from the image when it is used for pixel-level semantic segmentation. To overcome this issue, a fully convolutional network (FCN) has been developed where the dense fully connected part of CNN is replaced with deconvolution. This deconvolution layer takes the down-sampled feature maps from previous layers and up-samples them to original input image dimensions with dense pixel level segmentation [68]. FCN and its variants outclassed contemporary techniques in MedSeg tasks [69]. FCNs have been used for abdominal organ segmentation in 2D/3D images [70, 71], in the segmentation of the left ventricle (LV), right ventricle (RV), and myocardium in cardiac MRI [72]. Cascaded FCNs, in which each FCN utilizes predicted features of the previous FCN model to produce other features [73], has been used for liver lesion segmentation [74] and various other tasks such as detecting fetus boundaries in ultrasound images [75].

FCNs have their shortcomings. The responses at the final layers of FCNs are not localized enough to produce fine segmentation results. To solve this, Chen et al. [76] proposed DeepLab v1 based on a Visual Geometry Group (VGG)-16 network with a Conditional Random Field (CRF). The DeepLab system had better accuracy than previous models in localizing segmentation boundaries. DeepLab v2 [77], based on a more complex ResNet-101 and fully connected CRF solved some difficulties faced in DeepLab v1. DeepLab v2 was further improved to develop DeepLab v3 [78] and DeepLab v3 + [79] to refine segmentation results along the target objects' boundaries. Medical images can be of different modalities and resolutions but the input size for a typical FCN is fixed. Hence for different types and sizes (resolutions) of input images, FCN struggles. To overcome this issue, multi-stream FCN in conjunction with the multi-modality technique can be used to extract information from images of different resolutions and makes the system robust for different types of organ size and shape [80]. Based on the encoder-decoder structure of FCN, SegNet [81] is an improved model for pixel-level image segmentation. It continues to use VGG-16 in the encoder for feature maps but the novelty is that its decoder has lower-resolution up-sampled output. A larger pooling index is used in the decoder for up-sampling and then dense feature maps are produced from the convolution operation. A softmax classifier is used at the end to classify pixels (semantic segmentation) from the feature maps, restored to the original size. This way, spatial information of high frequency is maintained, a lesser number of parameters are required to be trained, and better edge characterization is achieved [4]. A 50 layers-deep fully convolutional residual network (FCRN) proposed by Yu et al. [82] was capable of pixel-level segmentation. It produced significantly better results than other models in skin lesions (melanoma) segmentation. Some other instances of FCN application in MedSeg tasks are discussed in section III.

2.3 U-Net

Based on the FCN architecture, U-Net is one of the most popular and researched DL structures for MedSeg tasks in images from a range of modalities [83]. Due to its success in various MedSeg applications, Isensee et al. [84] popularly stated that "a well-trained U-Net is hard to beat." It uses a novel deconvolution method [85] and skip connections between different layers in the encoder and decoder paths [73], as depicted in Fig. 1. The U-Net architecture overcomes the trade-off between context and localization accuracy. As given in Fig. 1, the U-Net model has a contracting route (a convolution network) and an expanding route (consisting deconvolution network) where the contracting route comprises convolution layers, ReLU activation function, and max-pooling layers, and the expanding route has deconvolution layers, each of which follows an up-sampling layer [80]. Skip connections between the layers of matching resolution of contractive and expansive paths provide feature maps from convolution layers to deconvolution layers, which proved very helpful in the segmentation task. Due to these novelties, the U-Net produces highly precise and faster segmentation results using a smaller number of labeled data. These properties are well-suited for MedSeg applications. Because of its popularity, U-Net and its various derivative structures have recently been developed for different types of MedSeg tasks and challenges.

For the better utilization of spatial information, a 3D U-Net model was developed for volumetric segmentation by Cicek et al. [86], consisting of a network similar to the 2D U-Net with all the 2D operations replaced by their 3D equivalents. Zhou et al. [87] proposed a deeply-supervised version of U-Net named U-Net++or nested U-Net in which the skip-connections of U-Net were replaced by dense nested pathways consisting of convolution blocks. The loss

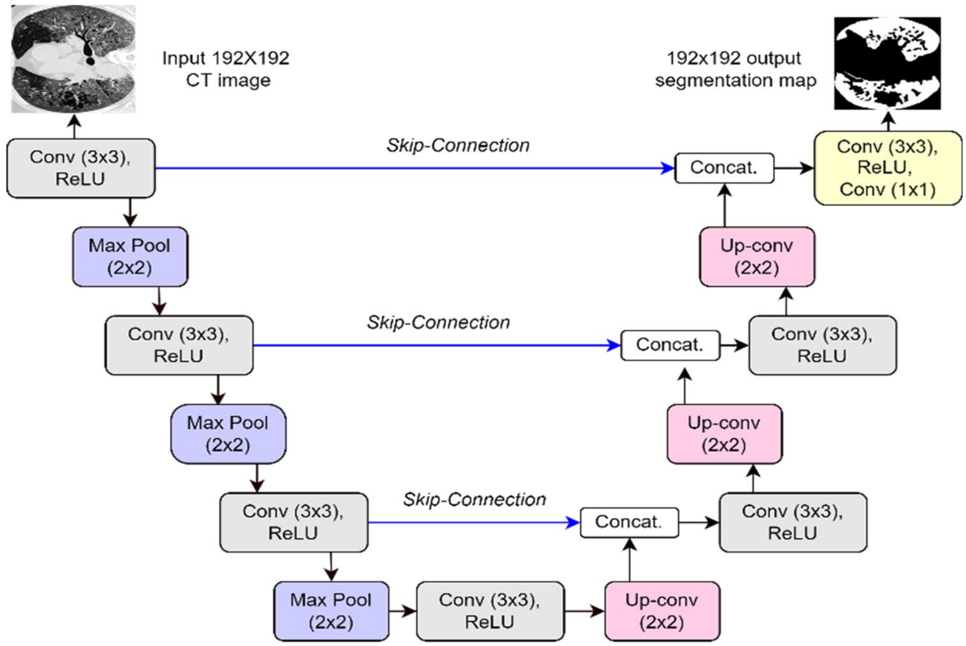

Fig. 1 U-Net model showing contracting and expanding paths with skip-connections in a U-shaped architecture

function, optimization function, and regularization method in U-Net++ have also been updated from the original U-Net. Authors used U-Net++on a wide range of MedSeg applications, such as nodule segmentation in chest CT images, liver segmentation in abdominal CT scans, nuclei segmentation in microscopy images, etc., outperforming other models. Isensee et al. [84] proposed 'No new-net' based on U-Net (also called nnU-Net) which, instead of focusing on architectural changes, focused on training schemes to produce much better segmentation results. They modified preprocessing steps, optimization function, loss function, and data augmentation techniques of the original U-Net. Oktay et al. [88] integrated attention mechanism with the U-Net so that the model could focus on the relevant areas of the image and ignore others.

The attention mechanism eliminates the requirement of object localization modules in the segmentation networks and improves the segmentation performance of the model. Alom et al. [89] combined U-Net with Recurrent Residual CNNs (R2U-Net), which improves the training by producing better feature representations and capturing contextual information through feature accumulation. Combining attention mechanism with R2U-Net, Zuo et al. [90] proposed a model (R2AU-Net) that can better capture context and focus on essential regions of the image. V-Net is another such famous derivative network of U-Net in which convolution kernels replace max-pooling for resolution reduction that ensures smaller memory footprints. In this structure, new objective and activation functions are used [91]. V-Net shows faster convergence and better performance on challenging medical segmentation datasets than other U-Net-based models. For highly variable medical images, Wang et al. [17] proposed AFD-UNet, an adaptive fully dense (AFD) network founded on U-Net for MedSeg. Through horizontal connections in a fully dense network, the AFD-UNet adaptively uses shallow and deep features from the network to learn pixel-level labels. The model can perform pixel-level transformations while retaining delicate edge information. The authors used this model to produce the best results in liver cancer segmentation in CT images.

To compare these U-Net-based models' performance in MedSeg, we have implemented them on the "Covid-19 lung infection CT segmentation dataset" [45], which has 98 annotated lung CT images from Covid-19 patients. All the models have been trained using the same labeled 50 lung CT images and tested on the remaining 48 images. A comparison of these models has been summarized in Table 1 with corresponding dice scores. A visual comparison of the segmentation results is shown in Fig. 2. Results show that the nnU-Net by Isensee et al. [84] gives the best segmentation performance on the given dataset. Since the model architecture of nnU-Net is the same as the baseline U-Net architecture, its remarkable performance can be attributed to the improvements in the training methodologies. This establishes that the baseline U-Net architecture is very robust and still relevant in medical image segmentation tasks.

2.4 Generative Adversarial Networks (GAN)

GANs are DL frameworks where two separate neural network structures are trained simultaneously. First is a generative network (G) that takes noise as input and generates fake images. The second one is a discriminative network (D) that determines whether the sample image came from training data (real image) or G (fake image), as shown in

Table 1 Comparison of U-Net-based models on the Covid-19 lung infection segmentation dataset

References	Model	Architectural/training approach	Dice score
Ronneberger et al. [84]	U-Net	Encoder-decoder-based structure with skip-connections	0.5569
Zhou et al. [87]	U-Net++	A deeply supervised U-Net with dense nested pathways	0.6962
Oktay et al. [88]	Attention-U-Net	U-Net with attention mechanism	0.6623
Isensee et al. [84]	nnU-Net	U-Net with a better training process	**0.7619**
Alom et al. [89]	R2U-Net	Combination of U-Net and Recurrent Residual CNNs	0.4757
Zuo et al. [90]	R2AU-Net	A combination of Attention-U-Net and R2U-Net	0.5411

Bold face indicates the best value among all

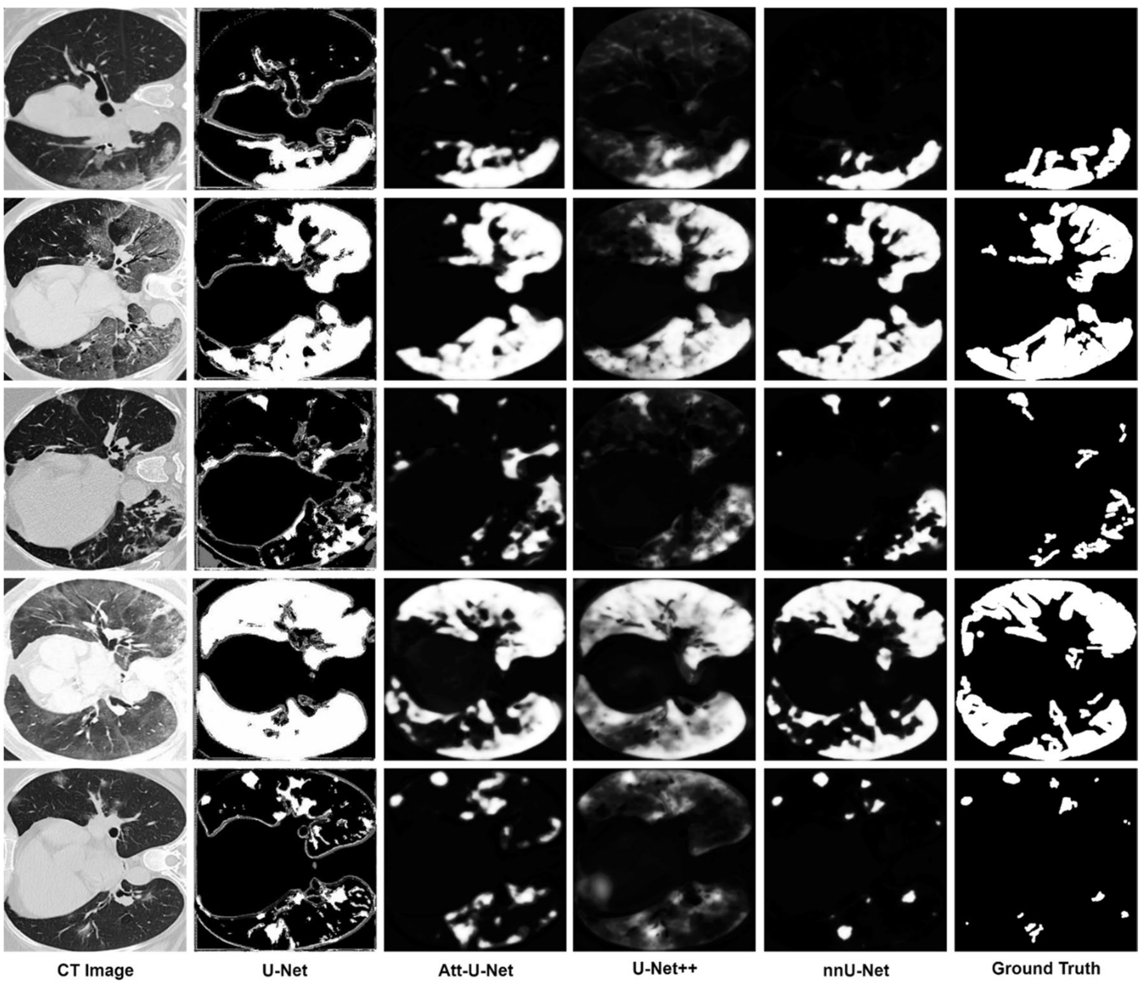

Fig. 2 Visual comparison of segmentation results of U-Net-based models on Covid-19 lung infection segmentation dataset

Fig. 3 [92]. In a way, there is competition between the two networks where G generates fake images and the adversarial network (discriminator) determines whether the input given to it is real or fake. GAN uses unsupervised learning to train its networks and learn abstract features in input images. Hence GANs are proved to be very useful where the availability of labeled data is scarce [93]. Luc et al. [94] used GAN for semantic segmentation, in which a CNN-based classifier is used as a generator which tries to generate segmentation maps close to the ground truth (GT) labels,and an adversarial network tries to discriminate between GT and generated images. GANs are being increasingly used in DL-based MedSeg applications. Xue et al. [95] used a GAN-based network for brain tumor segmentation. The network has two parts: a segmentor (generator) (S) and a critic (discriminator) (C). Here U-Net is

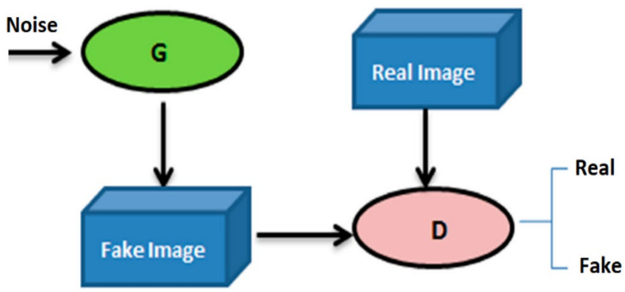

Fig. 3 A basic GAN structure where G represents the generator network and D represents discriminator network

used as the segmentor whose output and GT label maps are fed to the critic network.

A structure correction confrontation network (SCAN) was developed by Dai et al. [96] for lung and heart segmentation in Chest X-ray (CXR) images. In this structure, both the segmentor (generator) and the critic (discriminator) used FCN. Khosravan et al. [97] developed Projective Adversarial Network (PAN) for efficiently segmenting 3D medical images. This network has an FCN-based segmentor and two discriminative networks. Chang et al. [98] developed a secure and efficient GAN: 'Distributed Asynchronized Discriminator GAN (AsynDGAN)' made of a single generator and multiple discriminators for MedSeg. An encoder-decoder model is used for the generator. This model improves segmentation accuracy when compared to the other state-of-the-art models. In DL based MedSeg approaches such as image enhancements, domain adaptation, same-domain and cross-domain image synthesis, etc. GAN based architectures such as CycleGANs and conditional GANs are widely used. Some recent works with application of GAN based models in MedSeg tasks will be discussed in section III.

3 Deep Learning Approaches for Dataset Issues in MedSeg

The Major challenges in the field of MedSeg are the scarcity of high-quality annotated dataset and model generalizability. Getting reliable annotations for medical images is hard, time-taking and costly affair whereas distribution shifts in images from different sources create major generalizability problems. Researchers have followed different approaches to mitigate these problems. These approaches are generally focused on cost-effective and efficient annotation of medical images, augmentation of existing labeled datasets, utilizing external and unlabeled datasets, and implementation of several other techniques for improving the training of DL based segmentation network. Recent developments in these approaches are discussed in this section.

3.1 Active Learning and Interactive Segmentation

The process of annotating medical images is a hard, time-taking and costly process since it requires well-trained medical experts and high-end equipment. Due to this fact, it becomes extremely important to wisely select image samples for annotation from a large pool of unlabeled data and to speed up the whole process of annotation. Active learning (AL) is a cost-effective technique to judiciously select the most informative image samples from a large unlabeled dataset for annotation. In the AL process, the unlabeled images are fed to a segmentation model and using a selection mechanism dependent on the model output, next batch of image samples are selected for annotation. AL is generally an iterative process and is repeated until the best model performance is achieved. Active learning methods mostly differ by their sample selection mechanism Kim et al. [99] proposed a cascade of 3D U-Net with an AL scheme to efficiently train the model for kidney segmentation in abdominal CT scans with limited data. Initially a very small size training data is prepared using manual labeling and used to train the segmentation model. In the next stage, some additional data inputs are given to the trained model whose output is corrected manually. The annotated data produced in this step is used to enlarge original training dataset for subsequent training. This step is repeated until all the data is labeled and utilized. This method reduced labelling effort, saved time and improved dice score with each training step and gave superior performance compared to other 3D U-Net based models.

Wu et al. [100] proposed an AL framework for detecting COVID-19 from chest CT scans. This framework consisted of a U-Net for lung segmentation and considers sample diversity along with predicted loss as parts of its AL strategy. The diversity in CT volumes is calculated using a prediction confidence and a loss prediction network is used to predict the classification loss. The diversity and classification loss are merged to get the informativeness of a CT volume. Using this iterative hybrid strategy, most informative samples are selected from the unlabeled dataset pool and are labelled by the medical experts. Nath et al. [101] studied multiple methods of AL for MedSeg. They presented a query-by-committee using a joint optimizer for AL. The proposed AL framework with a U-Net based segmentation model is used for the segmentation of hippocampus MRI scans and CT scans of pancreas and tumors in which it attains full accuracy just by using about one fourth and half of each available dataset respectively. The framework is able to use highly uncertain samples based on model uncertainty through their increased frequency in the training set and reduces the overall annotation cost.

Zhao et al. [102] proposed a combined framework of AL and semi-supervised learning (SSL) for MedSeg. As a sample selection criterion, disagreement between intermediate features in the deeply supervised network is utilized for uncertainty measurement. Hence this criterion utilizes the information within the network and in each AL step, highly uncertain samples are given to strong labelers (oracle or expert annotator) and certain samples are given to weak labelers (pseudo-labels generated in SSL manner with dense conditional random field). The proposed framework with a lighter deeply supervised U-net as segmentation model was implemented on a skin lesion segmentation dataset having dermoscopy images and on a bone age X-ray dataset where it outperformed other baseline models. Most of the AL techniques are iterative in which informative samples are selected in multiple rounds for annotation by expert annotator. This requires more expert interaction with increased labor and computing costs. To overcome this issue, Jin et al. [103] aimed to select informative and diverse samples in just one step in a one-shot active learning scheme called 'Contrastive annotation', where contrastive self-supervised learning and farthest point sampling (diversity) based query strategy was used for selecting most informative samples. They performed skin lesion segmentation and chest X-ray segmentation using this framework and achieved better results from other state-of-the-art methods by a large margin.

Interactive segmentation can further complement the AL schemes in making annotation process cost-effective and fast. Interactive segmentation reduces the time required in annotating a selected image where the expert interactively corrects an initial segmentation mask produced by a model. The human experts review the initial segmentation mask produced by a model and interactively points out the errors in the mask. These interactions can be in the form of mouse-clicks, making bounding-boxes or scribblings to provide feedback to the segmentation model. Based on these feedbacks, the model regenerates segmentation mask and the whole process is repeated until desired results are obtained. Wang et al. [67] developed a DL based interactive 2D/3D MedSeg method that integrates CNNs and bounding-box and scribble-based user interactions. They also proposed an image-specific fine-tuning (with or without user-interactions) so that the CNNs better adopt to a specific test image. They implemented their framework for 2D fetal MRI for organ segmentation and 3D MRI brain tumor segmentation in which the model performed well on unseen objects and the model achieves faster and accurate results with fewer user interactions. Zhang et al. [104] applied interactive segmentation method for segmenting kidney tumors in CT images and prostate MRI.

Their method of interaction does not require bounding boxes or seed points but only need medical expert to click on the approximate center of the object before segmentation. Then based on this center point, a convolutional recurrent neural network with a gated memory propagation unit extracts different sequential patches in an inside-out manner. Finally, a U-Net is used for segmentation, and the whole process takes lesser time while improving the segmentation performance. Luo et al. [105] also employed a user mouse-click input based interactive segmentation method called that generalizes well to previously unseen objects for MedSeg. To encode these user interactions for adjusting a CNN for better initial segmentation, a context-aware and parameter-free method was developed. Table 2 summarizes these active learning and interactive segmentation-based methods for comparison.

3.2 Data Augmentation and Image Enhancement Techniques

Data augmentation (DA) can be an effective quantitative solution for model overfitting problem due to the scarcity of labeled data in MedSeg. DA techniques aim to enhance the size and quality of the available training datasets so that the DL model can be trained better [27]. DA methods can be broadly categorized into (a) traditional methods such as intensity transforms, spatial transforms and image mixings and (b) modern synthetic augmentation techniques. Image enhancement techniques focus on qualitative improvements in the datasets such as noise removal, contrast enhancement, improving image resolution, boundaries, etc. Here, we will focus on modern synthetic augmentation techniques and image enhancement techniques, which are in most cases, are based on the use of generative adversarial networks (GANs). Modern synthetic augmentation methods can be further divided in same-domain and cross-domain image synthesis methods. In same-domain synthesis, annotated data is synthesized directly in the target domain, whereas in cross-domain synthesis, some foreign data is projected into the target domain. In this section, we will be discussing same-domain synthesis methods as cross-domain synthesis is similar to domain-adaptation techniques which will be discussed in the next section.

Shin et al. [107] proposed a generative algorithm in which they used an image-to-image translation conditional GAN for synthetic generation of multi-parametric abnormal brain tumor MRI images using brain MRI datasets. The authors generated MRI images with high levels of variation in tumor sizes, location of tumor and generated images tumor placed in a healthy brain. Results showed that tumor segmentation models trained on both real and synthetic MR images performed significantly better without other forms of data augmentations applied. A Cycle-GAN transforms an image

Table 2 Summary of AL and Interactive segmentation methods

References	Object (Modality)	Segmentation model	Methodology
Kim et al. [99]	Kidney (Abdominal CT)	Cascaded 3D U-Net	CNN corrected segmentation in an iterative AL framework
Wu et al. [100]	Lung (Chest CT)	U-Net	Sample diversity and predicted loss as iterative AL strategy
Nath et al. [101]	Hippocampus (MRI) and pancreas & tumours (CT)	Modified U-Net	Query-by-committee with a joint optimizer as iterative AL scheme to select diverse and uncertain samples
Zhao et al. [102]	Skin lesion (Dermoscopy)	Deeply-supervised U-Net	Intermediate features disagreement for iterative AL
Jin et al. [103]	Skin lesion (Dermoscopy) and lung lesions (X-ray)	U-Net	Non-iterative one-shot AL scheme with contrastive self-supervised learning and diversity-based strategy
Wang et al. [67]	Fetal organs and brain tumour (2D/3D MRI)	P-Net [106]	Bounding-box and scribble-based interactive segmentation method with image specific fine-tuning
Zhang et al. [104]	Kidney tumours (CT) and Prostate (MR)	U-Net	Mouse-click based interactive segmentation method with inside-out patch extraction strategy
Luo et al. [105]	Placenta (fetal MRI), Spleen (abdomen CT), Brain tumour (MRI), etc	U-Net & 3D U-Net	Mouse-click based interactive segmentation method with parameter-free context-awareness

of one type to another. Gilbert et al. [108] used a Cycle-GAN to generate synthetic ultrasound images with corresponding labels from existing high quality annotations. These annotations were automatically produced from previously made anatomical models. The Cycle-GAN architecture is used in its default settings except that the generator network is replaced with a U-Net. Pseudo images generated from the anatomical structures work as input to the Cycle-GAN which is matched with the shape distribution of real ultrasound images to produce synthetic ultrasound images. In this way we get synthetic ultrasound images paired with high quality segmentation mask which is used to train a segmentation network (U-Net) for heart left ventricle (LV) and left atrium (LA) segmentation. The corresponding results are comparable to the network trained on real ultrasound images which proves that synthetically produced training data can work as real training data. This process could be applied to other MedSeg tasks in any image modality.

CNNs mainly focus on local features by design, which affects getting proper contextual information. To solve this problem, Dalmaz et al. [109] developed a medical image synthesis approach based on generative adversarial model, in which aggregated residual convolutional and vision transformer modules were implemented in generator to improves contextual information, while maintaining focus on localization. The model synthesizes superior quality brain MRI and multi-modal pelvic images compared to other CNN and transformer-based models. Thanbawita et al. [110] presented a framework for generating synthetic medical images along with paired masks using only a single real image-mask pair as input. The framework is a modified form of SinGAN [111] which is an unconditional generative model that uses only one natural image for training, to produce synthetic ones. The proposed framework was used to generate high quality synthetic polyp images paired with segmentation masks. A U-Net++ is trained on this synthetic data produces comparable and better results when it is trained on a large or small quantity of real data respectively. Li et al. [112] presented a multi-scale conditional GAN consisting a pyramid scheme for generating high-resolution histopathology images and their segmentation masks. The pyramid generation scheme allows synthesis of large-scale high-resolution histopathology images which includes richer context, improving segmentation. These synthetic images, paired with their masks were used to train a segmentation model in supervised and semi-supervised settings. The results demonstrated that the framework could be effectively used to augment limited pathology datasets and boost segmentation performance in semi-supervised settings.

Most of the medical image enhancement techniques mostly focus on denoising medical images, increasing their contrast and resolution (image super-resolution), improving object boundaries and overall quality of images. These techniques have proved to improve the performance of segmentation models for a wide range of MedSeg tasks. Tang et al. [113] used stacked GANs for CT image enhancement where the model removes noise from original CT images and produces high contrast and high-resolution images. Experiments demonstrate that the model enhanced images improve the performance of segmentation models better than the other enhancement methods compared. Prior anatomic knowledge of shape and location of organs in medical images are key to improve their segmentation. Oktay et al. [114] presented anatomically constrained CNNs that use anatomical prior knowledge for medical image enhancement and segmentation. Their approach used autoencoders and transfer-learning networks as regularisers for training CNNs on cardiac MRI dataset. Results demonstrate that CNNs performance improved by using anatomical priors, especially when the images are corrupted, less informative and inconsistent. Zhu et al. [115] proposed an arbitrary-scale image super-resolution (SR) method where they applied transfer learning approach to produce SR images with different scales for cardiac MRI and Covid-chest CT image modalities. The proposed method has the potential of improving performances in medical image enhancement, reconstruction and segmentation as pre- or post-processing step. Ma et al. [116] presented a relativistic average GAN to enhance the spatial resolution of lung CT and prostate MRI images in an SR technique where the generator is trained to produce high-resolution images using low-resolution input images. It is very difficult to get high resolution isotropic cardiac MRI cine images as it takes a long time to acquire while the patient holds breath. To solve this problem, Xia et al. [117] proposed a conditional GANs based SR method that synthesizes high-quality SR cardiac MRI images using visual and anatomical properties of input images in a transfer learning approach. These high-quality SR images are shown to improve cardiac segmentation tasks. A good aspect of this study was that the method didn't require high-resolution images or multiple low-resolution images to produce anatomically tenable SR images.

Uneven illumination and structural losses in medical images make it difficult for automatic analysis. To resolve this issue, Ma et al. [118] proposed structural and illumination constrained bi-directional GAN to enhance medical images in which better structural details and uniform illumination is achieved. The model learns to transfer the properties of high-quality images to low-quality ones under structure and illumination constraints. The produced high-quality images significantly improve nerve fiber segmentation in corneal confocal microscopy images. In a significantly different enhancement approach, You et al. [119] proposed a dual-rotation network, a special segmentation network, that enhances the feature maps (global, local, shallow and deep) of the network layers. To obtain a richer semantic information, the model learns and combines shallow, deep, local and global features and further enhances the feature maps through rotation, dilation and multi-scaling. Compared with other state-of-the-art segmentation models, the proposed model produces good segmentation results in abdominal CT-MRI organ segmentation and brain MRI tumor segmentation.

To summarize this section, we can say that recent same-domain image synthesis and image enhancement methods employ various types of GANs that include Cycle-GANs, Conditional GANs, and transformation networks. A number of recent works have focused on medical image super-resolution as enhancement and augmentation task. These methods have largely been helpful in mitigating the problem of data scarcity in MedSeg. The data augmentation and enhancement-based techniques are summarized in Table 3.

3.3 Transfer Learning and Domain Adaptation

To deal with the problem of dataset scarcity, it is possible to leverage external datasets by using techniques such as transfer learning (TL) and domain adaptation (DOA). In TL approaches, a model trained on external data (medical or non-medical) can be applied to the target medical image data. DOA techniques try to transform images from one modality (like CT, MRI, etc.) to other or from one image distribution to other in the same modality. In TL based approaches, the external data can either be used to pretrain the network (later fine-tuning by target training data) or for feature extraction. TL application in MedSeg tasks has been limited, partially due to the use of less deep networks which respond less to fine-tuning and also due to the 3D nature of most of the medical image modalities [5]. Performance of TL schemes are highly application and data (quantity and quality) dependent, however in most cases, TL based method reduce convergence time. When the segmentation task is more difficult and target labeled data is very small, only then we can observe significant improvements in model performance [120]. TL based approached are more common in medical image classification tasks but their application in MedSeg has been very limited. Liu et al. [121] proposed a model that transferred features from 2D network to 3D. The framework uses a 2D encoder that extracts features (parameters) from 2D slices which are transferred to the paired layers of 3D encoder of the proposed model. Then the decoder of the model is used to exploit the anisotropic 3D image context. The model gives better results when applied for liver tumor segmentation in CT volumes. Chen et al. [122] developed a heterogenous 3D model to extract 3D features from medical images. The pre-trained models were transferred to the tasks such as lung lesion segmentation and liver tumor segmentation having small training data, in which the model achieved faster training convergence time and improved segmentation performance. Niu et al. [123] proposed a distant domain TL approach for Covid-19 diagnosis. Their model could use somewhat unrelated (distant domain) data to produce better segmentation on lung CT scans, ultimately used for image classification and Covid-19 diagnostic.

DOA techniques try to align distribution shifted data from different domains, either by transforming data from one domain to another or by learning common features among them. DOA techniques have become popular because distribution shift between training and test data is a common problem in medical images due to the use of variety

Table 3 Summary of data augmentation and enhancement techniques

References	Object and modality	Task	Model
Shin et al. [107]	Brain tumour MRI	Image synthesis	Conditional GAN (pix2pix)
Gilbert et al. [108]	Cardiac Ultrasound	Image synthesis	Cycle-GAN with U-Net as generator
Dalmaz et al. [109]	Brain MRI and multi-modal pelvic	High-quality image synthesis	ResViT: A GAN with generator having residual convolutional and vision transformer
Thanbawita et al. [110]	Polyp images	Image synthesis	SinGAN-Seg: an unconditional generative model using single image as input
Li et al. [112]	Histopathology	High-res image synthesis	Multi-scale conditional GAN in a pyramid scheme
Tang et al. [113]	Lesion CT	Image enhancements	Stacked GANs (SGAN)
Oktay et al. [114]	Cardiac MRI	Image super-resolution	Anatomically constrained CNNs (ACNNs)
Zhu et al. [115]	Cardiac MR and Covid-Chest CT	Arbitrary scale image super-resolution	MIASSR: An arbitrary-scale SR method for medical images
Ma et al. [116]	Lung CT, Prostate MRI	Image super-resolution	Relativistic average GAN
Xia et al. [117]	Cardiac MRI	Image super-resolution	Conditional GAN in a transfer learning approach
Ma et al. [118]	Nerve fiber in corneal confocal microscopy	Image structure and illumination enhancement	StillGAN: structural and illumination constrained GAN
You et al. [119]	Abdominal CT-MR	Feature-maps enhancement	Dual-rotation network (DR-Net)

of scanners, scanning methods, and patient diversity [124]. As discussed earlier, DOA is similar to cross-domain synthesis where a foreign data is projected into target domain. Nie et al. [125] adopted generative adversarial approach to generate CT from MRI and 7-Tesla MRI from 3-Tesla MRI images. A fully convolutional network (FCN) was used as a generator and image gradient difference loss function is used to generate clearer target images. Auto-context model was applied to make the GAN context-aware. Results show that synthetic MRI images produce segmentation maps very similar to those produced by real MRI images. Use of GANs, especially cycle-consistent GANs (Cycle-GANs) is common in DOA methods for image reconstruction. Wang et al. [126] proposed a cycle-GAN based DOA method which is immune to domain specific noise and deformations and was used to synthesize multi-sequence brain MRI and cross-modality abdominal CT-MRI data. The model achieved good alignment between the source data and synthesized one. Luo et al. [127] developed a cross-modality MRI synthesis method that takes edge information into account. The proposed GAN based method which has three parts: (1) a target modality image synthesis network; (2) an edge generation network and (3) an adversarial network to distinguish between synthesized and real images in target domain. The method was applied to brain tumor MRI segmentation and proved to be superior to other models. When the target medical image data is unlabeled i.e., ground-truth masks are not available, unsupervised domain adaptation (UDA) methods have proved to be very helpful in segmentation. UDA methods utilize labeled images from other domains for training a target domain-centric model. Liu et al. [128] presented a scheme for cardiac segmentation on different modalities (CT & MRI). This cross-modal technique incorporated a UDA scheme with an attention structure to focus on the important regions of the images from multiple modalities. Hence the segmentation scheme learns the relevant cross-modal features to predict segmentation masks regardless of the input image domain. This modal performed better than other cross-modal segmentation schemes. Wu and Zhuang [129] proposed a UDA method in which two networks based on variational auto-encoders were trained on labeled data from source domain and unlabeled data from target domain to drive the latent features from both domains towards a common distribution and applied a regularization for variational approximation. The method is applied to cross-modality CT-MRI heart segmentation and cross-sequence cardiac MRI segmentation tasks and achieved best results.

Xie et al. [130] proposed a UDA method for MedSeg, which is based on disentanglement learning which breaks an image into domain-invariant anatomical features and domain-specific features. Domain-invariant anatomical features are further utilized with a shape-constraint for domain adaptation. To improve the segmentation performance in target domain, self-training methods are also used. The method showed best results in cardiac, abdominal and brain segmentation. Wang and Yao et al. [131] proposed many-to-many mapping in a UDA scheme and introduced 3D segmentation framework in DOA to include semantic knowledge from different depth levels in a medical image volume. The model was applied to cross-modality brain structures and abdominal organs segmentation in MRI-CT volumes.

Liu et al. [132] also proposed a UDA method where category-aware contrastive learning scheme is used for feature alignment rather than matching global marginal distributions in a category-agnostic way in images from different domains. The model is proved to be among the best in whole heart segmentation task in cross-modal CT-MRI images. To address the problem of learning from noisy labels and distribution shifted data from different sources, Liu et al. [133] proposed an UDA method that learns common features from different domain images, excluding the noise. The model uses two peer adversarial networks for the above-mentioned tasks. This method was applied for optic cup and optic disc segmentation in retinal fundus images, as well as spinal cord gray matter segmentation in MR images. The method outperformed previous models in cleaning noisy labels from source domain and robustly adapting to unlabeled target domain to produce best predictions for it. Zero-shot learning (ZSL) helps DL models to recognize unseen object classes, hence it can be a used in the segmentation of unseen target structures by utilizing cross-modality features. Bian et al. [134] proposed a ZSL based framework for the segmentation of cross-modality (MRI-CT) cardiac and abdominal datasets. The method could segment unseen objects in target images by training on labeled source data and using cross-modality info. The TL and DOA methods are summarized in Table 4.

3.4 Self-Supervised and Semi-supervised Learning

Unlabeled data can be utilized along with labeled data to better train segmentation models. It can be done in two ways: (1) Self-supervised learning (SFSL) in which the unlabeled data is used to pre-train a segmentation model for some substitute tasks like classification, denoising, enhancement, regression, etc. and then updated using the available labeled dataset. (2) Semi-supervised learning (SSL) in which unlabeled data is used with the labeled data to train a model for segmentation task. SFSL techniques differ by the substitute tasks for which the model's weights are initialized using unlabeled data. As a better

Table 4 Summary of Transfer learning and Domain adaptation-based methods

References	Object and Modality	Model	Remarks
Liu et al. [121]	Liver tumour in CT	AH-Net: 3D anisotropic hybrid network	Transfer learning
Chen et al. [122]	Lung lesion and liver tumour in CT	Med3D: a heterogeneous 3D network	Transfer learning
Niu et al. [123]	Covid-19 lung lesions in CT	DDTL: Distant Domain Transfer Learning	Transfer learning
Nie et al. [125]	Brain MRI	GAN with FCN as generator	Cross-domain data transformation
Wang et al. [126]	Brain MRI and abdominal CT-MRI	DiCyc: Deformation invariant Cycle-GAN	Cross-domain data transformations immune to domain-specific deformations
Luo et al. [127]	Brain tumour MRI	EP_IMF-GAN: Edge preserving, iterative multi-scale fusion based GAN	Edge-preserving cross-modality MRI synthesis
Liu et al. [128]	Cardiac CT-MRI	UMDA-SNA-SFCNN: multi-UDA, spatial neural attention, symmetric-FCN	Cross-modal segmentation framework based on UDA and attention mechanism
Wu and Zhuang [129]	Heart CT-MR and cardiac MRI	Two networks based on variational auto-encoders (VAEs)	Cross-modal UDA method based on VAEs
Xie et al. [130]	Cardiac MRI-CT, abdominal CT, and brain tumour MRI	DLaST: UDA with disentangled learning and self-training	UDA scheme where factored domain-invariant features are used for adaptation
Liu et al. [133]	Optic cup and disc in retinal fundus images and spinal-cord MRI	S-CUDA: Self-cleansing unsupervised domain adaptation	Two peer adversarial networks trained for learning common features in cross-domain images excluding the noise
Yao et al. [131]	Brain and abdominal organs in MR-CT	3D UDA scheme with a self-attention module	3D cross-modality segmentation
Bian et al. [134]	Cardiac and abdominal objects in MRI-CT images	Zero-shot-learning based DOA	Cross-modality feature extraction for domain adaptation
Liu et al. [132]	Cardiac CT-MRI	MPSCL: margin preserving self-paced contrastive learning	Contrastive learning-based UDA for cross-modal segmentation

alternative to TL for MedSeg, SFSL models are pre-trained on the same-domain medical images, rather than those from different domains. Chen et al. [135] applied SFSL technique for segmentation of multi-modal brain MRI in which context-restoration was used as surrogate self-supervision task for pretraining the model. Context restoration task enables the CNN based model to learn important features of an image which can be useful in subsequent medical image analysis tasks such as segmentation. Zheng et al. [136] proposed a task agnostic hierarchical SFSL technique which pre-trains an encoder-decoder based network on heterogenous unlabeled medical image datasets for subsequent multiple segmentation tasks. The authors collected multi-modality datasets to pre-train the model for learning task-agnostic hierarchical features. Fang et al. [137] proposed a multi-domain SFSL based segmentation structure in which multi-modal brain tumor MRI images from multiple datasets are input to a multi-encoder network to extract multi-domain features which are combined by a hybrid attention fusion module to learn hybrid features of multi-modal images. These hybrid features, learned in a self-supervised way, ultimately improves the brain tumor segmentation task in the given modality. Ouyang et al. [138] proposed a SFSL few-shot based scheme for MedSeg where in an unsupervised feature learning mechanism, the model uses super-pixel based pseudo-annotations for self-supervision. The model was used for multi-organ segmentation in CT-MRI images and cardiac segmentation in MRI and produces better results than conventional few-shot segmentation methods which required manual labeling. Unsupervised learning approaches can identify abnormal regions but not the target disease [24–26]. But an SSL based model can detect regions associated with the target disease by learning from the available small size labeled dataset along with a large size unlabeled dataset. This explains the recent interests in the use of SSL methods generally [139] and in MedSeg too [5, 37]. SSL approaches can be further divided into techniques that generate pseudo-labels and those that don't. In pseudo-label generating methods, pseudo-annotations are generated for the unlabeled data in an iterative manner and then both the labeled and pseudo-labeled data are used to train the segmentation model. If the domain of the unlabeled data is different from the target labeled data, then the method is commonly known as unsupervised domain adaptation (UDA), about which we have already discussed in the previous section. In this section, we will discuss SSL based approaches with pseudo-labels and without pseudo-labels having same-domain labeled and unlabeled data.

Nie et al. [140] proposed a region-attention based SSL framework where in addition to the segmentation network, a confidence network is also included which helps in efficiently choosing a part of unlabeled data for segmentation training.

The confidence network is essentially an FCN-based discriminator which, through adversarial learning, generates local confidence maps. This method was applied to segment prostate, bladder and rectum in pelvic MRI, where it significantly outperforms other models in accuracy and robustness. Fan et al. [141] proposed an SSL scheme where pseudo-labels were generated for a large pool of unlabeled data in a progressive random selection scheme to augment the available labeled data. The semi-supervised version of their model outperformed the supervised only version and other state-of-the-art models by a significant margin in segmenting lung infections in CT images of Covid-19 patients. Chaitanya et al. [142] proposed a technique where a generator produces augmentation data, which is improved using labeled and unlabeled data in an SSL framework. The model outperforms standard augmentation and SSL based model in segmenting pancreas CT and prostate MRI images. Wu et al. [143] used a structure having a single shared encoder and multiple slightly different decoders to compute uncertainties by comparing decoder outputs. High uncertainty outputs denote challenging regions for segmentation in the image where a consistency constraint is applied to force the model to produce similar results. The whole network is trained in semi-supervised way and produces superior performance in MedSeg. For covid-19 lung infection segmentation with fewer annotated images, Wang et al. [144] developed an iterative SSL based few-shot learning framework in which pseudo-labels are generated for a large amount of unannotated data and added to the small annotated training dataset for a more effective training. Re-weighing and confidence values are used to improve the trustworthiness of pseudo-labels. The model outperformed state-of-the-art models in segmenting infected lung regions in chest CT images. Luo et al. [145] proposed a pyramid consistency regularization approach for SSL in which the model learns from the unlabeled data by encouraging the multi-scale pyramid predictions to be alike. An uncertainty correction mechanism is also implemented so that the model could learn only from the reliable areas of the images. The model is applied to segment Nasopharyngeal carcinoma (NPC) and brain tumor in MRI, and pancreas in abdominal CT scans where it achieves better or comparable results compared to other SSL based methods. These studies prove that SSL based approaches are better for medical applications where a large pool of same-domain unlabeled data is available along with the limited annotated data. The SFSL and SSL based techniques are summarized in Table 5.

Table 5 Summary of Semi- and Self-supervised learning-based methods

References	Modality	Method	Remarks
Chen et al. [135]	Multi-modal Brain MRI	SFSL	Context-restoration as substitute task
Zheng et al. [136]	Multi-organ CT-MRI	HSSL: Hierarchical SFSL	Task-agnostic pre-training on multi-modal datasets
Fang et al. [137]	Multi-modal Brain tumour MRI	Multi-modal SFSL	Multi-modal feature fusion
Ouyang et al. [138]	Multi-organ CT-MRI, Cardiac MRI	SSL-ALPNet: an SFSL few-shot method	Super-pixel based pseudo-annotations used for self-supervision
Nie et al. [140]	Prostate, bladder and rectum in pelvic MRI	ASDNet: Attention based SSL	Pseudo-label free SSL using local confidence maps
Fan et al. [141]	Lung Infection in CT images	Semi-Inf-Net: SSL based segmentation network	Pseudo-label-based SSL method using attention mechanism and edge info
Chaitanya et al. [142]	Prostate MRI and pancreas CT	DA based on SSL	A task-driven SSL based DA method
Wang et al. [144]	Lung infection in chest CT	SSA-Net: spatial self-attention network with SSL few-shot framework	A pseudo-label-based iterative semi-supervised few-shot framework
Luo et al.[145]	Nasopharyngeal carcinoma and brain tumour in MRI, pancreas in abdominal CT	URPC: uncertainty rectified pyramid consistency for SSL	A pyramid consistency regularization approach for SSL with uncertainty correction

4 Challenges and Opportunities

In this study, we have broadly covered developments related to deep learning architectures and approaches that aim to overcome some of the major dataset related challenges faced in DL based automated MedSeg. These challenges include scarcity of quality data for training due to difficulties in getting accurate annotations and privacy concerns, distribution shift among datasets from different image modalities and acquisition sources, images with noise, low contrast, and fuzzy boundaries etc. In the architectural development section, we first discussed conventional CNNs which required large amount of annotated data for training and suffered from problems like loss of spatial information from the image when it is used for pixel-level segmentation. Encoder-decoder architecture based FCNs were developed to overcome these issues by modifying the standard CNN architecture. FCNs had their own challenges such as the responses at the final layers of FCNs are not localized enough to produce fine segmentation results. To overcome those challenges, some FCN architecture was modified into U-shaped structures with skip-connections called U-Nets. U-Nets have since become very popular and are some of the most-researched architectures in the field of DL based MedSeg. Several U-Net based architectures such as U-Net++ and attention-U-Net have been developed to improve MedSeg performance. Finally, an increasing use of generative adversarial models is observed in MedSeg for tasks such as image enhancements, data augmentation, domain adaptation, etc. using GAN based architectures such as CycleGANs and conditional GANs.

Along with these architectural developments, several approaches focusing mostly on the non-architectural aspects of deep learning have also been developing to address the major challenges faced in MedSeg. These approaches include (a) active learning and interactive segmentation which focus on cost-effective, fast and efficient annotation of medical images, (b) data augmentation and enhancement techniques which try to enhance the quantity and quality of training data, (c) transfer learning and domain adaptation techniques which try to use foreign datasets for improving model training, and (d) self-supervised and semi-supervised learning which utilize larger pool of unlabeled datasets for improving model performances.

From this study, we have come to know that the scarcity of high-quality labeled data and distribution shift among datasets, remain some of the biggest challenges in the path of creating a highly robust, accurate and generalizable model for automatic MedSeg applications. It is for this reason, recently we have seen a lot of works in the fields of (1) advanced image enhancement techniques, especially the image super-resolution techniques, that aim to enhance the overall quality of images, (2) unsupervised domain adaptation techniques that utilizes datasets from different domains to pre-train the segmentation model in an unsupervised manner, and (3) semi-supervised learning based techniques that aims to augment limited available training data with a larger pool of unlabeled data. Deep learning based MedSeg still has a lot of challenges and scope of work in the fields we have mentioned above.

5 Conclusion

In the past few years, deep learning-based methods have rapidly become state-of-the-art in medical image segmentation tasks. But these methods face some major dataset related challenges that include image quality issues, labeled data scarcity, distribution shift among datasets, model generalizability, etc. In this paper, we have reviewed recent developments in deep learning architectures and approaches that aim to resolve these challenges. We covered CNNs, FCN, U-Net, GAN and their derivative models applied in medical image segmentation tasks. We also implemented some of the popular U-Net based models on a Covid-19 lung infection dataset to compare their segmentation performances. We studied some key approaches that focus on the non-architectural aspects of deep learning such as active learning, data augmentation, domain adaptation, semi-supervised learning, etc., aiming to provide ingenious solutions for major challenges in the field. Even with these developments, model generalizability and dataset related problems remain formidable challenges in this field. It is for this reason, among other approaches; super-resolution-based image enhancement, unsupervised domain adaptation, and semi-supervised learning-based techniques have been hot areas of research recently with a large scope of work remaining in these fields.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

References

1. Bhandari AK, Singh A, Kumar IV (2019) Spatial context energy curve-based multilevel 3-D Otsu algorithm for image segmentation. IEEE Trans Syst Man Cybern 51(5):2760–2773
2. Bhandari AK, Kumar IV, Srinivas K (2019) Cuttlefish algorithm-based multilevel 3-D Otsu function for color image segmentation. IEEE Trans Instrum Meas 69(5):1871–1880
3. Kumar S, Bhandari AK, Raj A, Swaraj K (2021) Triple clipped histogram-based medical image enhancement using spatial frequency. IEEE Trans Nanobiosci 20(3):278–286
4. Sushanki S, Bhandari AK, Singh AK (2023) A review on computational methods for breast cancer detection in ultrasound images using multi-image modalities. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-023-10015-0
5. Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X (2020) Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med Image Anal 63:101693. https://doi.org/10.1016/j.media.2020.101693
6. Budd S, Robinson EC, Kainz B (2021) A survey on active learning and human-in-the-loop deep learning for medical image analysis. Med Image Anal 71:102062. https://doi.org/10.1016/j.media.2021.102062
7. Li C et al (2013) A likelihood and local constraint level set model for liver tumor segmentation from CT volumes. IEEE Trans Biomed Eng 60(10):2967–2977. https://doi.org/10.1109/TBME.2013.2267212
8. Upadhyay AK, Bhandari AK (2023) Semi-supervised modified-U-Net for lung infection image segmentation. IEEE Trans Radiat Plasma Med Sci. https://doi.org/10.1109/TRPMS.2023.3272209
9. Held K, Kops ER, Krause BJ, Wells WM, Kikinis R, Müller-Gärtner HW (1997) Markov random field segmentation of brain MR images. IEEE Trans Med Imaging 16(6):878–886. https://doi.org/10.1109/42.650883
10. Lalonde M, Beaulieu M, Gagnon L (2001) Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. IEEE Trans Med Imaging 20(11):1194–1200. https://doi.org/10.1109/42.963823
11. Pandey SK, Bhandari AK, Singh H (2022) A transfer learning based deep learning model to diagnose covid-19 CT scan images. Heal Technol 12(4):845–866
12. Tsai A et al (2003) A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans Med Imaging 22(2):137–154. https://doi.org/10.1109/TMI.2002.808355
13. Preity AKB, Shahnawazuddin S (2023) Automated computationally intelligent methods for ocular vessel segmentation and disease detection: a review. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-023-09940-x
14. Pandey SK, Bhandari AK (2023) A systematic review of modern approaches in healthcare systems for lung cancer detection and classification. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-023-09940-x
15. Verma PR, Bhandari AK (2023) Role of deep learning in classification of brain mri images for prediction of disorders: a survey of emerging trends. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-023-09967-0
16. Latif J, Xiao C, Imran A and Tu S (2019) Medical imaging using machine learning and deep learning algorithms: A review. In: 2019 2nd Int. Conf Comput Math Eng Technol iCoMET 2019. doi: https://doi.org/10.1109/ICOMET.2019.8673502.
17. Wang EK, Chen CM, Hassan MM, Almogren A (2020) A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain. Futur Gener Comput Syst 108:135–144. https://doi.org/10.1016/j.future.2020.02.054
18. Wang S, Yang DM, Rong R, Zhan X, Xiao G (2019) Pathology image analysis using segmentation deep learning algorithms. Am J Pathol 189(9):1686–1698. https://doi.org/10.1016/j.ajpath.2019.05.007
19. Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: overview, challenges and the future. Lect Notes Comput Vis Biomech 26:323–350. https://doi.org/10.1007/978-3-319-65981-7_12
20. Zhang Y, Gorriz JM, Dong Z (2021) Deep learning in medical image analysis. J Imaging. https://doi.org/10.3390/jimaging7040074
21. Ker J, Wang L, Rao J, Lim T (2017) Deep learning applications in medical image analysis. IEEE Access 6:9375–9379. https://doi.org/10.1109/ACCESS.2017.2788044
22. Lundervold AS, Lundervold A (2019) An overview of deep learning in medical imaging focusing on MRI. Z Med Phys 29(2):102–127. https://doi.org/10.1016/j.zemedi.2018.11.002
23. López-Linares Román K, García Ocaña MI, Let Urzelai N, González Ballester MÁ, Macía Oliver I (2020) Medical image segmentation using deep learning. Intell Syst Ref Libr 171:17–31. https://doi.org/10.1007/978-3-030-32606-7_2
24. Zhou K et al. (2020) Sparse-Gan: sparsity-constrained generative adversarial network for anomaly detection in retinal OCT image. In: Proc - Int Symp Biomed Imaging. doi: https://doi.org/10.1109/ISBI45749.2020.9098374.
25. Chalapathy R and Chawla S (2019) Deep learning for anomaly detection: a survey. Available: http://arxiv.org/abs/1901.03407
26. Schlegl T, Seeböck P, Waldstein SM, Schmidt-Erfurth U, Langs G (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. Lect Notes Comput Sci. https://doi.org/10.1007/978-3-319-59050-9_12

27. Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data. https://doi.org/10.1186/s40537-019-0197-0
28. Hesamian MH, Jia W, He X, Kennedy P (2019) Deep learning techniques for medical image segmentation: achievements and challenges. J Digit Imaging 32(4):582–596. https://doi.org/10.1007/s10278-019-00227-x
29. Rizwan I, Neubert J (2020) Deep learning approaches to biomedical image segmentation. Informatics Med Unlocked 18:100297. https://doi.org/10.1016/j.imu.2020.100297
30. Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J (2020) A survey on U-shaped networks in medical image segmentations. Neurocomputing 409:244–258. https://doi.org/10.1016/j.neucom.2020.05.070
31. Siddique N, Paheding S, Elkin CP, Devabhaktuni V (2021) U-net and its variants for medical image segmentation: a review of theory and applications. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3086020
32. Yin XX, Sun L, Fu Y, Lu R, Zhang Y (2022) U-Net-based medical image segmentation. J Healthc Eng. https://doi.org/10.1155/2022/4189781
33. Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. Med Image Anal. https://doi.org/10.1016/j.media.2019.101552
34. Jeong JJ, Tariq A, Adejumo T, Trivedi H, Gichoya JW, Banerjee I (2022) Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation. J Digit Imaging 35(2):137–152. https://doi.org/10.1007/s10278-021-00556-w
35. Xun S et al (2022) Generative adversarial networks in medical image segmentation: a review. Comput Biol Med 140:105063. https://doi.org/10.1016/j.compbiomed.2021.105063
36. Asgari Taghanaki S, Abhishek K, Cohen JP, Cohen-Adad J, Hamarneh G (2021) Deep semantic segmentation of natural and medical images: a review. Springer, Heidelberg. https://doi.org/10.1007/s10462-020-09854-1
37. Cheplygina V, de Bruijne M, Pluim JPW (2019) Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Med Image Anal 54:280–296. https://doi.org/10.1016/j.media.2019.03.009
38. Bohlender S, Oksuz I, Mukhopadhyay A (2021) A survey on shape-constraint deep learning for medical image segmentation. IEEE Rev Biomed Eng 3333:1–17. https://doi.org/10.1109/RBME.2021.3136343
39. Xie X, Niu J, Liu X, Chen Z, Tang S, Yu S (2021) A survey on incorporating domain knowledge into deep learning for medical image analysis. Med Image Anal 69:101985. https://doi.org/10.1016/j.media.2021.101985
40. Zhou T, Ruan S, Canu S (2019) "A review: deep learning for medical image segmentation using multi-modality fusion. Array. https://doi.org/10.1016/j.array.2019.100004
41. Devunooru S, Alsadoon A, Chandana PWC, Beg A (2021) Deep learning neural networks for medical image segmentation of brain tumours for diagnosis: a recent review and taxonomy. J Ambient Intell Humaniz Comput 12(1):455–483. https://doi.org/10.1007/s12652-020-01998-w
42. Wadhwa A, Bhardwaj A, Singh Verma V (2019) A review on brain tumor segmentation of MRI images. Magn Reson Imaging 61:247–259. https://doi.org/10.1016/j.mri.2019.05.043
43. Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X (2021) A review of deep learning based methods for medical image multi-organ segmentation. Phys Medica 85:107–122. https://doi.org/10.1016/j.ejmp.2021.05.003
44. Shi F et al (2021) Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE Rev Biomed Eng 14:4–15. https://doi.org/10.1109/RBME.2020.2987975
45. COVID-19 CT Segmentation Dataset (2020). https://medicalsegmentation.com/covid19/ Accessed 15 Aug 2022
46. Bengio Y, Courville A, Vincent P (2013) Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 35(8):1798–1828. https://doi.org/10.1109/TPAMI.2013.50
47. Wang G (2016) A perspective on deep imaging. IEEE Access 4:8914–8924. https://doi.org/10.1109/ACCESS.2016.2624938
48. Suzuki K (2017) Overview of deep learning in medical imaging. Radiol Phys Technol 10(3):257–273. https://doi.org/10.1007/s12194-017-0406-5
49. Albawi S, Mohammed TA and Al-Zawi S (2018) Understanding of a convolutional neural network. In: Proc 2017 Int Conf Eng Technol ICET 2017. doi: https://doi.org/10.1109/ICEngTechnol.2017.8308186.
50. Agarap AF (2018) Deep learning using rectified linear units (ReLU). Available: http://arxiv.org/abs/1803.08375
51. Kingma DP and Ba JL (2015) Adam: a method for stochastic optimization. In: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.
52. Commandeur F et al (2018) Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast CT. IEEE Trans Med Imaging 37(8):1835–1846. https://doi.org/10.1109/TMI.2018.2804799
53. Zhang W et al (2015) Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. Neuroimage 108:214–224. https://doi.org/10.1016/j.neuroimage.2014.12.061
54. Roth HR et al (2014) A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. Lect Notes Comput Sci 8673:520–527. https://doi.org/10.1007/978-3-319-10404-1_65
55. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M (2013) Deep feature learning for knee cartilage segmentation using a Triplanar convolutional neural network. Lect Notes Comput Sci 8150:246–253. https://doi.org/10.1007/978-3-642-40763-5_31
56. Moeskops P et al (2016) Deep learning for multi-task medical image segmentation in multiple modalities. Lect Notes Comput Sci 9901:478–486. https://doi.org/10.1007/978-3-319-46723-8_55
57. Gupta P and Malhi AK (2018) Using deep learning to enhance head and neck cancer diagnosis and classification. In: 2018 IEEE Int Conf Syst Comput Autom Networking, ICSCA 2018. doi: https://doi.org/10.1109/ICSCAN.2018.8541142.
58. Dou Q et al (2017) 3D deeply supervised network for automated segmentation of volumetric medical images. Med Image Anal 41:40–54. https://doi.org/10.1016/j.media.2017.05.001
59. Kleesiek J et al (2016) Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. Neuroimage 129:460–469. https://doi.org/10.1016/j.neuroimage.2016.01.024
60. Jakab M, Stevuliak M and Benesova W (2020) Multi-modal brain tumor segmentation utilizing convolutional neural networks. doi: https://doi.org/10.1117/12.2557599.
61. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2016.90
62. Chen H, Dou Q, Yu L, Qin J, Heng PA (2018) VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. Neuroimage 170:446–455. https://doi.org/10.1016/j.neuroimage.2017.04.041
63. Zhao F, Xie X (2013) An overview of interactive medical image segmentation. Ann Br Mach Vis Assoc 2013(7):1–22

64. Rajchl M et al (2017) DeepCut: object segmentation from bounding box annotations using convolutional neural networks. IEEE Trans Med Imaging 36(2):674–683. https://doi.org/10.1109/TMI.2016.2621185
65. Xu N, Price B, Cohen S, Yang J, Huang T (2016) Deep interactive object selection. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2016.47
66. Ribeiro HL and Gonzaga A (2006) Hand image segmentation in video sequence by GMM: a comparative analysis. In: Brazilian Symp Comput Graph Image Process. doi: https://doi.org/10.1109/SIBGRAPI.2006.23.
67. Wang G et al (2018) Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans Med Imaging 37(7):1562–1573. https://doi.org/10.1109/TMI.2018.2791721
68. Roth HR et al. (2018) Deep learning and its application to medical image segmentation. vol 36, pp. 63–71. doi: https://doi.org/10.11409/mit.36.63.
69. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651. https://doi.org/10.1109/TPAMI.2016.2572683
70. Zhou X, Takayama R, Wang S, Hara T, Fujita H (2017) Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Med Phys 44(10):5221–5233. https://doi.org/10.1002/mp.12480
71. Hu P, Wu F, Peng J, Bao Y, Chen F, Kong D (2017) Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. Int J Comput Assist Radiol Surg 12(3):399–411. https://doi.org/10.1007/s11548-016-1501-5
72. Wong KKL, Fortino G, Abbott D (2020) Deep learning-based cardiovascular image diagnosis: a promising challenge. Futur Gener Comput Syst 110:802–811. https://doi.org/10.1016/j.future.2019.09.047
73. Christ PF et al (2016) Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. Lect Notes Comput Sci 9901:415–423. https://doi.org/10.1007/978-3-319-46723-8_48
74. Christ PF et al. (2017) Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. Available: http://arxiv.org/abs/1702.05970
75. Wu L, Xin Y, Li S, Wang T, Heng PA and Ni D (2017) Cascaded Fully Convolutional Networks for automatic prenatal ultrasound image segmentation. In: Proc - Int Symp Biomed Imaging. doi: https://doi.org/10.1109/ISBI.2017.7950607.
76. Chen L, Kokkinos I, Murphy K and Yuille AL (2015) S EMANTIC I MAGE S EGMENTATION WITH D EEP C ON.
77. Chen L, Papandreou G, Member S, Kokkinos I, Murphy K, Yuille AL (2018) DeepLab : semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848
78. Chen LC, Papandreou G, Schroff F and Adam H (2017) Rethinking atrous convolution for semantic image segmentation. Available: http://arxiv.org/abs/1706.05587
79. Chen L, Zhu Y, Papandreou G, Schroff F and Aug CV Encoder-decoder with atrous separable convolution for semantic image segmentation.
80. Zeng G and Zheng G (2018) Multi-stream 3D FCN with multi-scale deep supervision for multi-modality isointense infant brain MR image segmentation Guodong Zeng and Guoyan Zheng Institute for Surgical Technology & Biomechanics. University of Bern. Bern.
81. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615
82. Yu L, Chen H, Dou Q, Qin J, Heng PA (2017) Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans Med Imaging 36(4):994–1004. https://doi.org/10.1109/TMI.2016.2642839
83. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4
84. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2019) No new-net. Lect Notes Comput Sci 11384:234–244. https://doi.org/10.1007/978-3-030-11726-9_21
85. Suah FBM (2017) Preparation and characterization of a novel Co(II) optode based on polymer inclusion membrane. Anal Chem Res 12:40–46. https://doi.org/10.1016/j.ancr.2017.02.001
86. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. Lect Notes Comput Sci 9901:424–432. https://doi.org/10.1007/978-3-319-46723-8_49
87. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J (2018) Unet++: a nested u-net architecture for medical image segmentation. Lect Notes Comput Sci 11045:3–11. https://doi.org/10.1007/978-3-030-00889-5_1
88. Oktay O et al (2018) Attention U-Net: learning where to look for the pancreas. Available: http://arxiv.org/abs/1804.03999
89. Alom VKA, Zahangir Md, Hasan M, Yakopcic C, Taha TM (2020) Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. Lect Notes Comput Sci 12085:207–219. https://doi.org/10.1007/978-3-030-47436-2_16
90. Zuo Q, Chen S, Wang Z (2021) R2AU-Net: attention recurrent residual convolutional neural network for multimodal medical image segmentation. Secur Commun Netw. https://doi.org/10.1155/2021/6625688
91. Milletari F, Navab N and Ahmadi SA (2016) V-Net: fully convolutional neural networks for volumetric medical image segmentation.In: Proc - 2016 4th Int Conf 3D Vision, 3DV 2016. doi: https://doi.org/10.1109/3DV.2016.79.
92. Goodfellow I et al (2020) Generative adversarial networks. Commun ACM 63(11):139–144. https://doi.org/10.1145/3422622
93. Majurski M et al. (2019) Cell image segmentation using generative adversarial networks, transfer learning, and augmentations. In: IEEE Comput Soc Conf Comput Vis Pattern Recognit Work. doi: https://doi.org/10.1109/CVPRW.2019.00145.
94. Luc P, Couprie C, Chintala S and Verbeek J (2016) Semantic segmentation using adversarial networks. Available: http://arxiv.org/abs/1611.08408
95. Xue Y, Xu T, Zhang H, Long LR, Huang X (2018) SegAN: adversarial network with multi-scale L 1 loss for medical image segmentation. Neuroinformatics 16(3–4):383–392. https://doi.org/10.1007/s12021-018-9377-x
96. W. Dai, N. Dong, Z. Wang, X. Liang, H. Zhang, and E. P. Xing, “Scan: Structure correcting adversarial network for organ segmentation in chest x-rays,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11045 LNCS, pp. 263–273, 2018, doi: https://doi.org/10.1007/978-3-030-00889-5_30.
97. Khosravan N, Mortazi A, Wallace M, Bagci U (2019) PAN: projective adversarial network for medical image segmentation. Lect Notes Comput Sci 11769:68–76. https://doi.org/10.1007/978-3-030-32226-7_8
98. Chang Q et al. (2020) Synthetic learning: learn from distributed asynchronized discriminator GaN without sharing medical image data. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. doi: https://doi.org/10.1109/CVPR42600.2020.01387.

99. Kim T et al (2020) Active learning for accuracy enhancement of semantic segmentation with CNN-corrected label curations: evaluation on kidney segmentation in abdominal CT. Sci Rep 10(1):1–7. https://doi.org/10.1038/s41598-019-57242-9
100. Wu X, Chen C, Zhong M, Wang J, Shi J (2021) COVID-AL: the diagnosis of COVID-19 with deep active learning. Med Image Anal 68:101913. https://doi.org/10.1016/j.media.2020.101913
101. Nath V, Yang D, Landman BA, Xu D, Roth HR (2021) Diminishing uncertainty within the training pool: active learning for medical image segmentation. IEEE Trans Med Imaging 40(10):2534–2547. https://doi.org/10.1109/TMI.2020.3048055
102. Zhao Z, Zeng Z, Xu K, Chen C, Guan C (2021) DSAL: deeply supervised active learning from strong and weak labelers for biomedical image segmentation. IEEE J Biomed Heal Informatics 25(10):3744–3751. https://doi.org/10.1109/JBHI.2021.3052320
103. Jin Q, Yuan M, Qiao Q, Song Z (2022) One-shot active learning for image segmentation via contrastive learning and diversity-based sampling. Knowledge-Based Syst 241:108278. https://doi.org/10.1016/j.knosys.2022.108278
104. Zhang J et al (2021) Interactive medical image segmentation via a point-based interaction. Artif Intell Med 111:101998. https://doi.org/10.1016/j.artmed.2020.101998
105. Luo X et al (2021) MIDeepSeg: Minimally interactive segmentation of unseen objects from medical images using deep learning. Med Image Anal 72:102102. https://doi.org/10.1016/j.media.2021.102102
106. Wang G et al (2019) DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. IEEE Trans Pattern Anal Mach Intell 41(7):1559–1572. https://doi.org/10.1109/TPAMI.2018.2840695
107. H. S. B et al Medical image synthesis for data augmentation and anonymization using generative adversarial networks. vol 1, pp. 1–11, doi: https://doi.org/10.1007/978-3-030-00536-8.
108. Gilbert A, Marciniak M, Rodero C, Lamata P, Samset E, Mcleod K (2021) Generating synthetic labeled data from existing anatomical models : an example with echocardiography segmentation. IEEE Trans Med Imaging 40(10):2783–2794. https://doi.org/10.1109/TMI.2021.3051806
109. Dalmaz O, Yurt M, Tolga C (2022) ResViT : residual vision transformers for multi-modal medical image synthesis. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2022.3167808
110. Id VT, Salehi P, Sheshkal SA, Hicks SA, Id AR (2022) SinGAN-Seg : synthetic training data generation for medical image segmentation. PLoS ONE. https://doi.org/10.1371/journal.pone.0267976
111. Shaham TR, Dekel T and Michaeli T (2019) SinGAN: learning a generative model from a single natural image. In: Proc IEEE Int Conf Comput Vis. doi: https://doi.org/10.1109/ICCV.2019.00467.
112. Li W, Li J, Polson J, Wang Z, Speier W, Arnold C (2022) High resolution histopathology image generation and segmentation through adversarial training. Med Image Anal 75:102251. https://doi.org/10.1016/j.media.2021.102251
113. Tang Y et al (2018) CT image enhancement using stacked generative adversarial networks and transfer learning for lesion segmentation improvement. Lect Notes Comput Sci 11046:46–54. https://doi.org/10.1007/978-3-030-00919-9_6
114. Oktay O et al (2018) Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans Med Imaging 37(2):384–395. https://doi.org/10.1109/TMI.2017.2743464
115. Zhu J, Tan C, Yang J, Heart N, Lio P (2021) Arbitrary scale super-resolution for medical images. Int J Neural Syst. https://doi.org/10.1142/S0129065721500374
116. Ma Y et al (2021) Medical image super-resolution using a relativistic average generative adversarial network. Nucl Instruments Methods Phys Res 992:165053. https://doi.org/10.1016/j.nima.2021.165053
117. Xia Y, Ravikumar N, Greenwood JP, Neubauer S, Petersen SE, Frangi AF (2021) Super-resolution of cardiac MR cine imaging using conditional GANs and unsupervised transfer learning. Med Image Anal 71:102037. https://doi.org/10.1016/j.media.2021.102037
118. Ma Y, Liu J, Member S, Liu Y, Fu H, Member S (2021) Structure and illumination constrained GAN for medical image enhancement. IEEE Trans Med Imaging 40(12):3955–3967
119. You H, Yu L, Tian S, Cai W (2022) DR-Net: dual-rotation network with feature map enhancement for medical image segmentation. Complex Intell Syst 8(1):611–623. https://doi.org/10.1007/s40747-021-00525-4
120. Karimi D, Warfield SK, Gholipour A (2021) Transfer learning in medical image segmentation: new insights from analysis of the dynamics of model parameters and learned representations. Artif Intell Med 116:102078. https://doi.org/10.1016/j.artmed.2021.102078
121. Liu S et al (2018) 3D anisotropic hybrid network: transferring convolutional features from 2D images to 3D anisotropic volumes. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-00934-2_94
122. Chen S, Ma K and Zheng Y (2019) Med3D: transfer learning for 3D medical image analysis. Available: http://arxiv.org/abs/1904.00625
123. Niu S, Liu M, Liu Y, Wang J, Song H (2021) Distant domain transfer learning for medical imaging. IEEE J Biomed Heal Informatics 25(10):3784–3793. https://doi.org/10.1016/B978-0-12-374979-6.00016-2
124. Guan H, Liu M (2022) Domain adaptation for medical image analysis: a survey. IEEE Trans Biomed Eng 69(3):1173–1185. https://doi.org/10.1109/TBME.2021.3117407
125. Trullo R, Lian J, Wang L, Petitjean C, Ruan S (2018) Medical image synthesis with deep convolutional adversarial networks. IEEE Trans Biomed Eng 65(12):2720–2730
126. Wang C et al (2021) DiCyc: GAN-based deformation invariant cross-domain information fusion for medical image synthesis. Inf Fusion. https://doi.org/10.1016/j.inffus.2020.10.015
127. Luo Y et al (2021) Edge-preserving MRI image synthesis via adversarial network with iterative multi-scale fusion. Neurocomputing 452:63–77. https://doi.org/10.1016/j.neucom.2021.04.060
128. Liu J et al (2021) Automated cardiac segmentation of cross-modal medical images using unsupervised multi-domain adaptation and spatial neural attention structure. Med Image Anal 72:102135. https://doi.org/10.1016/j.media.2021.102135
129. Wu F, Zhuang X (2021) Unsupervised domain adaptation with variational approximation for cardiac segmentation. IEEE Trans Med Imaging 40(12):3555–3567. https://doi.org/10.1007/s00180-021-01063-w
130. Xie Q et al (2022) Unsupervised domain adaptation for medical image segmentation by disentanglement learning and self-training. IEEE Trans Med Imaging. https://doi.org/10.1109/TMI.2022.3192303
131. Yao K et al (2022) A novel 3D unsupervised domain adaptation framework for cross-modality medical image segmentation. IEEE J Biomed Heal Informatics. https://doi.org/10.1109/jbhi.2022.3162118
132. Liu Z, Zhu Z, Zheng S, Liu Y, Zhou J, Zhao Y (2022) Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. IEEE J Biomed Heal Informatics 26(2):638–647. https://doi.org/10.1109/JBHI.2022.3140853
133. Liu L, Zhang Z, Li S, Ma K, Zheng Y (2021) S-CUDA: Self-cleansing unsupervised domain adaptation for medical image

segmentation. Med Image Anal 74:102214. https://doi.org/10.1016/j.media.2021.102214
134. Bian C, Yuan C, Ma K, Yu S, Wei D, Zheng Y (2022) Domain adaptation meets zero-shot learning: an annotation-efficient approach to multi-modality medical image segmentation. IEEE Trans Med Imaging 41(5):1043–1056. https://doi.org/10.1109/TMI.2021.3131245
135. Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D (2019) Self-supervised learning for medical image analysis using image context restoration. Med Image Anal 58:101539. https://doi.org/10.1016/j.media.2019.101539
136. Zheng H et al (2021) Hierarchical self-supervised learning for medical image segmentation based on multi-domain data aggregation. Lect Notes Comput Sci 12901:622–632. https://doi.org/10.1007/978-3-030-87193-2_59
137. Fang F, Yao Y, Zhou T, Xie G, Lu J (2021) Self-supervised multi-modal hybrid fusion network for brain tumor segmentation. IEEE J Biomed Heal Informatics 2194:1–11. https://doi.org/10.1109/JBHI.2021.3109301
138. Ouyang DR, Cheng, Biffi C, Chen C, Kart T, Qiu H (2022) Self-supervised Learning for Few-shot medical image segmentation In: IEEE Trans Med Imaging. Available: http://arxiv.org/abs/2003.08462
139. van Engelen JE, Hoos HH (2020) A survey on semi-supervised learning. Mach Learn 109(2):373–440. https://doi.org/10.1007/s10994-019-05855-6
140. Nie D, Gao Y, Wang L, Shen D (2018) ASDNet: attention based semi-supervised deep networks for medical image segmentation. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-00937-3_43
141. Fan DP et al (2020) Inf-Net: automatic COVID-19 lung infection segmentation from CT images. IEEE Trans Med Imaging 39(8):2626–2637. https://doi.org/10.1109/TMI.2020.2996645
142. Chaitanya K et al (2021) Semi-supervised task-driven data augmentation for medical image segmentation. Med Image Anal 68:101934. https://doi.org/10.1016/j.media.2020.101934
143. Wu Y et al (2022) Mutual consistency learning for semi-supervised medical image segmentation. Med Image Anal 81:102530. https://doi.org/10.1016/j.media.2022.102530
144. Wang X et al (2022) SSA-Net: spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning. Med Image Anal 79:102459. https://doi.org/10.1016/j.media.2022.102459
145. Luo X et al (2022) Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency. Med Image Anal 80:102517. https://doi.org/10.1016/j.media.2022.102517

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.